Caltech

Http Data Federations in CMS Project

J. Balcas, I. Legrand, D. Kcira, T. Hendricks

Open Computing Resource Workshop

CERN, June 21 2016

HTTP Data Federation Project in CMS

Caltech

- **Goals:**

  (1) Demonstrate ability of using HTTP data federations
  in a manner analogous to today's AAA infrastructure.

  (2) Improve CMSSW's HTTP support in case we see increased use of HTTP as a transport protocol

HTTP Data Federation Project in CMS (1)

Caltech

❑ Software Deliverables

- Implement a libdavix-based StorageFactory plugin in CMSSW to replace the current fork/exec of curl for HTTP access
- Develop extensions to the Xrootd [HTTP implementation] to replace any identified missing functionality (e.g. client-based monitoring IDs)
- As necessary, optimize the HTTP implementation within Xrootd to improve any performance deficiencies observed in the tests below
- As necessary, develop patches to better integrate HTTP-over-Xrootd with the OSG distribution (e.g. Authorization)
- (OPTIONAL) Develop asynchronous APIs to libdavix analogous to the current XrootD ones (so we can develop a multistream client for HTTP)

- **Plugin is almost ready (~1 more week):**
  - Supports HTTP(s), WebDAV and Amazon S3 protocols (http, https, dav, davs, s3)
  - Allows to control plugin through environment variables (Davix_DEBUG, X509_*, /tmp/x509up_u%d, AwsAuthKeys)
  - Vector operations (Partial reads, multi-range, single range) and get full file content to I/O buffer.

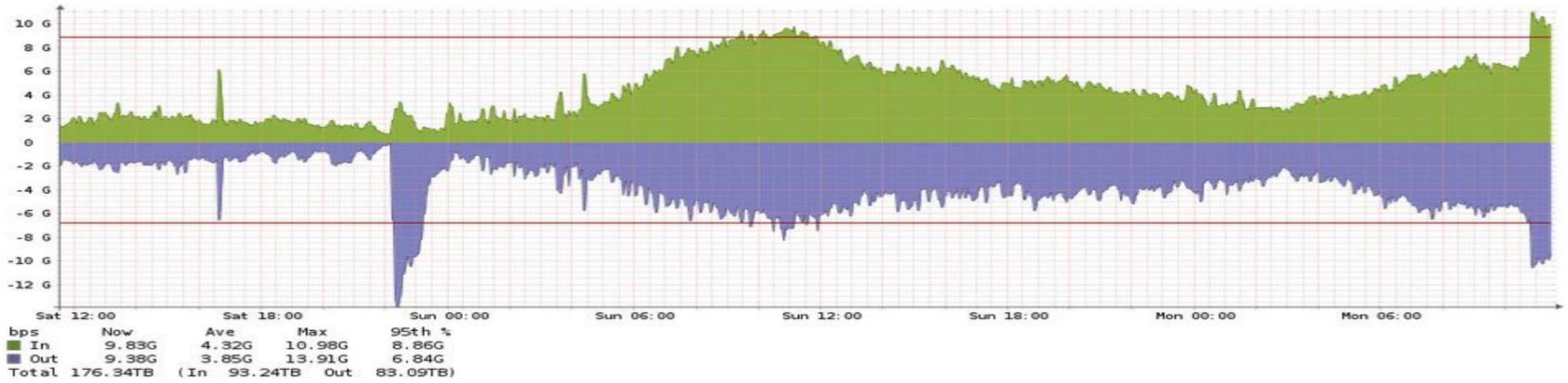
HTTP Data Federation Project in CMS (2)

Caltech

✓ **Deliverables: Operational**

- Improve support infrastructure at Caltech for data federations; Aim for a system that can saturate 80 Gbps
- Enable HTTP support at the US redirector and at the Caltech site. In collaboration with AAA and Caltech T2

- **Status update:**
  - Started from 1 xrootd server at 10Gbps and moved to 8 active xrootd servers total: **4 production + 4 testbed**
  - Set up production testbed, added an SSD to transfer-8 to test local FS performance
  - Bought new **100Gbps switch between CACR and Lauritsen.** Was commissioned 2 weeks ago.
  - Moving IPAC servers to Lauritsen => heavy data replication took place and link from CACR to Lauritsen was upgraded to 40GE. **No apparent congestions or large spikes were noticed.**



HTTP Data Federation Project in CMS (3)

Caltech

❑ Deliverables: Knowledge Base

- Document configuration & deployment of HTTP-over-Xrootd at Caltech
- Measure operational performance of a single Xrootd daemon at exporting w/HTTP
- Measure comparative performance in using the Caltech T2 via standard Xrootd versus HTTP through the AAA redirector
- Measure the comparative performance in using 'curl'-based StorageFactory plugin versus the libdavix one above
- Characterize the performance of using Xrootd for HTTP caching versus squid

---

- **Working on understanding Hadoop performance on US CMS Sites:**
  - Caltech (dd - 650Mbps, xrdcp – 350Mbps), Wisconsin (xrdcp – 350Mbps), Nebraska (asked to do same tests). All tests are done locally from HDFS to /dev/null.
- **Other deliverable tests will be done when libdavix plugin is ready.**

Next steps forward improving

Caltech

- **In collaboration with XrootD developers we are working to better understand and improve XrootD operations and some particular aspects, like:**
  - Large number of unused ports
  - Fallback decision procedure
  - Extend current XrootD monitoring and measure tcp throughput between all sites, do continuous ping to obtain the RTT and it's variation, use tracepath for routing information between all sites, more details [1]
- **An overflow mechanism was introduced by Brian as classad generation**
  - ➢ If a workflow subtask of selected type is idling too much (1:5 idle/running ratio) jobs are offloaded to neighbouring sites
  - ✓ Helps to rebalance the load
  - Would benefit from a dynamic mapping of what sites are in good XrootD terms
  - Make more decisions checking the number of currently running overflowed jobs and their success rate;
  - ✓ With same classad rules memory requirements are rounded for same group of jobs, if they are over-requesting memory in time. (API for Unified is available through cms-gwmsmon.cern.ch)

[1] CMS Production, XrootD and ML services https://indico.cern.ch/event/527183/